CLAIMS

What is claimed is:

1. A method for performing parallel operations on a pair of objects in a system that includes a plurality of nodes to enable using an increased degree of parallelism, the method comprising the computer-implemented steps of:

distributing first-phase partition-pairs of a parallel partition-wise operation on the pair of objects among the plurality of nodes;

at a particular node of said plurality of nodes, performing the steps of:

partitioning the one or more first-phase partition-pairs distributed to the particular node to produce a set of second-phase partition-pairs; and

assigning each second-phase partition-pair from the set of second-phase partition-pairs to a separate slave process.

2. The method of Claim 1 wherein the step of assigning each second-phase partition-pair from the set of second-phase partition-pairs to a separate slave process is performed by assigning each second-phase partition-pair from the set of second-phase partition-pairs to a separate slave process within said particular node.

3. The method of Claim 1, wherein the parallel partition-wise operation is a parallel full partition-wise operation.

4. The method of Claim 1, wherein the parallel partition-wise operation is a parallel partial partition-wise operation.

5. The method of Claim 1, wherein the step of partitioning the one or more first-phase partition-pairs and the step of assigning second-phase partition-pairs are performed at each node that has multiple slave processes available for participating in said parallel partition-wise operation.

6. The method of Claim 1, further comprising the steps of:

determining whether a total number of slave processes available for participating in a second parallel partition-wise operation has a particular logical relationship to a number of first-phase partition-pairs of the second parallel partition-wise operation;

if the total number of slave processes available for participating in the second parallel partition-wise operation has said particular logical relationship to the number of first-phase partition-pairs of the second parallel partition-wise operation, then at said particular node performing the steps of:

partitioning the one or more first-phase partition-pairs distributed to the particular node for the second parallel partition-wise operation to produce a set of second-phase partition-pairs; and

assigning the second-phase partition-pairs from the set of second-phase partition-pairs to slave processes within the particular node to cause the number of slave processes participating in said second parallel partition-wise operation on said particular node to be greater than the number of first-phase partition-pairs that were distributed to said particular node;

if the total number of slave processes available for participating in the second parallel partition-wise operation does not have the particular logical relationship to the number of first-phase partition-pairs of the second parallel partition-wise operation, then distributing said first-phase partition-pairs to slave processes without performing second-phase partitioning.

7. The method of Claim 6, wherein the step of partitioning the one or more first-phase partition-pairs and the step of assigning the second-phase partition-pairs are performed at each node that has more slave processes available for participating in said second parallel partition-wise operation than the number of first-phase partition-pairs that are distributed to the node.

8. The method of Claim 6, wherein the total number of slave processes available for participating in the second parallel partition-wise operation has the particular logical relationship to the number of first-phase partition-pairs of the second parallel partition-wise operation if the total number of slave processes available for participating in the second parallel partition-wise operation is greater than the number of first-phase partition-pairs of the second parallel partition-wise operation.

9. The method of Claim 6, wherein the total number of slave processes available for participating in the second parallel partition-wise operation has the particular logical relationship to the number of first-phase partition-pairs of the second parallel partition-wise operation if the total number of slave processes available for

participating in the second parallel partition-wise operation is at least an order of magnitude greater than the number of first-phase partition-pairs of the second parallel partition-wise operation.

10. The method of Claim 1, wherein the step of distributing first-phase partition pairs is performed based, at least in part, on node affinity with respect to the one or more first-phase partition-pairs of the parallel partition-wise operation and availability of slave processes for performing the parallel partition-wise operation.

11. A computer-readable medium carrying instructions for performing parallel operations on a pair of objects in a system that includes a plurality of nodes to enable using an increased degree of parallelism, the instructions comprising instructions for performing the computer-implemented steps of:

distributing first-phase partition-pairs of a parallel partition-wise operation on the pair of objects among the plurality of nodes;

at a particular node of said plurality of nodes, performing the steps of:

partitioning the one or more first-phase partition-pairs distributed to the particular node to produce a set of second-phase partition-pairs; and

assigning each second-phase partition-pair from the set of second-phase partition-pairs to a separate slave process.

12. The computer-readable medium of Claim 11 wherein the step of assigning each second-phase partition-pair from the set of second-phase partition-pairs to a separate

slave process is performed by assigning each second-phase partition-pair from the set of second-phase partition-pairs to a separate slave process within said particular node.

13. The computer-readable medium of Claim 11, wherein the parallel partition-wise operation is a parallel full partition-wise operation.

14. The computer-readable medium of Claim 11, wherein the parallel partition-wise operation is a parallel partial partition-wise operation.

15. The computer-readable medium of Claim 11, wherein the step of partitioning the one or more first-phase partition-pairs and the step of assigning second-phase partition-pairs are performed at each node that has multiple slave processes available for participating in said parallel partition-wise operation.

16. The computer-readable medium of Claim 11, further comprising instructions for performing the steps of:

determining whether a total number of slave processes available for participating in a second parallel partition-wise operation has a particular logical relationship to a number of first-phase partition-pairs of the second parallel partition-wise operation;

if the total number of slave processes available for participating in the second parallel partition-wise operation has said particular logical relationship to the number of first-phase partition-pairs of the second parallel partition-wise operation, then at said particular node performing the steps of:

partitioning the one or more first-phase partition-pairs distributed to the particular node for the second parallel partition-wise operation to produce a set of second-phase partition-pairs; and

assigning the second-phase partition-pairs from the set of second-phase partition-pairs to slave processes within the particular node to cause the number of slave processes participating in said second parallel partition-wise operation on said particular node to be greater than the number of first-phase partition-pairs that were distributed to said particular node;

if the total number of slave processes available for participating in the second parallel partition-wise operation does not have the particular logical relationship to the number of first-phase partition-pairs of the second parallel partition-wise operation, then distributing said first-phase partition-pairs to slave processes without performing second-phase partitioning.

17. The computer-readable medium of Claim 16, wherein the step of partitioning the one or more first-phase partition-pairs and the step of assigning the second-phase partition-pairs are performed at each node that has more slave processes available for participating in said second parallel partition-wise operation than the number of first-phase partition-pairs that are distributed to the node.

18. The computer-readable medium of Claim 16, wherein the total number of slave processes available for participating in the second parallel partition-wise operation has

the particular logical relationship to the number of first-phase partition-pairs of the second parallel partition-wise operation if the total number of slave processes available for participating in the second parallel partition-wise operation is greater than the number of first-phase partition-pairs of the second parallel partition-wise operation.

19. The computer-readable medium of Claim 16, wherein the total number of slave processes available for participating in the second parallel partition-wise operation has the particular logical relationship to the number of first-phase partition-pairs of the second parallel partition-wise operation if the total number of slave processes available for participating in the second parallel partition-wise operation is at least an order of magnitude greater than the number of first-phase partition-pairs of the second parallel partition-wise operation.

20. The computer-readable medium of Claim 11, wherein the step of distributing first-phase partition pairs is performed based, at least in part, on node affinity with respect to the one or more first-phase partition-pairs of the parallel partition-wise operation and availability of slave processes for performing the parallel partition-wise operation.

21. A method for performing parallel operations on a pair of objects including a source object and a target object in a broadcasting operation, the method comprising the computer-implemented steps of:

mapping each tuple from a source object to a corresponding static partition of a plurality of static partitions of the target object;

distributing the static partitions among the plurality of nodes according to a node distribution criteria; and

assigning each static partition to a slave process; and

broadcasting each tuple only to a group of slave processes assigned to the static partition to which the tuple is mapped.

22. The method of Claim 21, wherein the node distribution criteria includes node affinity with respect to the one or more first-phase partition-pairs of the parallel partition-wise operation and availability of slave processes for performing the parallel partition-wise operation.

23. A computer-readable medium carrying instructions for performing parallel operations on a pair of objects including a source object and a target object in a broadcasting operation, the instructions comprising instructions for performing the computer-implemented steps of:

mapping each tuple from a source object to a corresponding static partition of a plurality of static partitions of the target object;

distributing the static partitions among the plurality of nodes according to a node distribution criteria; and

assigning each static partition to a slave process; and

broadcasting each tuple only to a group of slave processes assigned to the static partition to which the tuple is mapped.

24. The computer-readable medium of Claim 23, wherein the node distribution criteria includes node affinity with respect to the one or more first-phase partition-pairs of the parallel partition-wise operation and availability of slave processes for performing the parallel partition-wise operation.